JOURNAL OF MULTIVARIATE ANALYSIS 3, 93–101 (1973)

Interaction Information and Markov Chain

SEIJI TAKANO

Department of Mathematics, Tokyo Institute of Technology, Tokyo, Japan

Communicated by P. K. Sen

The problem of multivariate information analysis is considered. First, the interaction information in each dimension is defined analogously according to McGill [4] and then applied to Markov chains. The property of interaction information zero deeply relates to a certain class of weakly dependent random variables. For homogeneous, recurrent Markov chains with m states, $m \geqslant n \geqslant 3$, the zero criterion of n-dimensional interaction information is achieved only by $(n - 2)$-dependent Markov chains, which are generated by some nilpotent matrices. Further for Gaussian Markov chains, it gives the decomposition rule of the variables into mutually correlated subchains.

1. INTRODUCTION

McGill [4] has introduced a multivariate interaction information for the exact analysis of the mutual informations. We define this more precisely from the information analysing standpoint, using the general formulations in the information theory given by Dobrushin [1], and then applying it to Markov chains, where interaction information plays a very important role.

We show that under the Markov condition, n-dimensional interaction information of any, differently selected n random variables takes negative and positive values corresponding to the dimensional degree to be odd and even, respectively, and especially, its value becomes zero if and only if the first and the last elements of n variables of the chain are mutually independent. Next, applying to a simple, homogeneous Markov chain, taking only finitely many recurrent states with the number m, it is proved that n-dimensional interaction information of any, differently selected n random variables is zero only when

Received July 24, 1972; revised October 30, 1972.

AMS 1970 subject classification numbers: Primary 94A15, 62H30, 60J10; Secondary 60J05, 60J15, 60K35.

Key words and phrases: Interaction information; Markov chain; M-dependence; multivariate information analysis; information theory.

this Markov chain follows the $(n-2)$-dependence, provided $m \geqslant n \geqslant 3$. On the other hand, we show another result for the Gaussian Markov chain. It asserts that for some dimension, an interaction information is zero if and only if this chain can be decomposed into some mutually correlated subchains. As these results, we have arrived at such a weak dependence where the variables are connected by Markov relations with some independent pairs to be contained.

2. Basic Definitions and Fundamental Results

Let $\{\xi_j\}$ $(j = 1, 2,...)$ be random variables, taking values in the measurable spaces (X_j, S_{x_j}) $(j = 1, 2,...)$, where X_j is a state space and S_{x_j}, its σ-algebra. Denote $P_{\xi_j}(\cdot)$, $P_{\xi_j\xi_k}(\cdot, \cdot)$ and $P_{\xi_j\times\xi_k}(\cdot, \cdot)$, the induced measures by the random variables ξ_j and (ξ_j, ξ_k), where the last measure is defined by

$$P_{\xi_j\times\xi_k}(\cdot, \cdot) = P_{\xi_j}(\cdot)\, P_{\xi_k}(\cdot).$$

Suppose that $P_{\xi_j\xi_k}$ is absolutely continuous with respect to $P_{\xi_j\times\xi_k}$. Then a mutual information between ξ_j and ξ_k is given [1]

$$I(\xi_j, \xi_k) = \int_{X_j\times X_k} \log \frac{P_{\xi_j\xi_k}(dx_j\, dx_k)}{P_{\xi_j\times\xi_k}(dx_j\, dx_k)}\, P_{\xi_j\xi_k}(dx_j\, dx_k). \tag{1}$$

Also, a conditional information between ξ_i and ξ_k, given ξ_j is defined

$$EI(\xi_i, \xi_k \mid \xi_j) = \int_{X_i\times X_j\times X_k} \log \frac{P_{\xi_i\xi_j\xi_k}(dx_i\, dx_j\, dx_k)}{\bar{P}_{\xi_i\times\xi_k|\xi_j}(dx_i\, dx_j\, dx_k)}\, P_{\xi_i\xi_j\xi_k}(dx_i\, dx_j\, dx_k), \tag{2}$$

which is equivalent to

$$I((\xi_i, \xi_j), \xi_k) - I(\xi_j, \xi_k) \qquad \text{or} \qquad I(\xi_i, (\xi_j, \xi_k)) - I(\xi_i, \xi_j),$$

where $\bar{P}_{\xi_i\times\xi_k|\xi_j}$ is the Dobrushin's measure [1, 5], and the equivalent form is derived from the formula of Dobrushin and Kolmogorov [1, 5].

Definition 1. A three dimensional interaction information of the three random variables ξ_i, ξ_j, and ξ_k is defined

$$J(\xi_i\xi_j\xi_k) = EI(\xi_i, \xi_k \mid \xi_j) - I(\xi_i, \xi_k). \tag{3}$$

If, for simplicity, we denote ξ_j by j only, it is written as

$$J(ijk) = EI(i, k \mid j) - I(i, k). \tag{3'}$$

This definition is the same as that of McGill [4]. We must take care that this representation is symmetric for each variable. In fact, we easily see that

$$\begin{aligned} J(ijk) &= EI(j, k \mid i) - I(j, k) \\ &= EI(k, i \mid j) - I(k, i) \\ &= EI(i, j \mid k) - I(i, j). \end{aligned} \tag{4}$$

By the way, it is well known that the information amount becomes infinite if the measure $P_{\xi_i \xi_j}$ is not absolutely continuous with respect to the measure $P_{\xi_i \times \xi_j}$ [7], so that hereafter, we assume the absolute continuity of the measure $P_{\xi_1 \xi_2 \ldots \xi_n}$ with respect to the measure $P_{\xi_1 \times \xi_2 \times \ldots \times \xi_n}$. The next lemma is fundamental.

LEMMA 1. (Dobrushin [1, 5]). *$EI(i, k \mid j) = 0$ occurs if and only if the variables ξ_i, ξ_j, and ξ_k form a simple Markov chain in this order. This is also true for the reversed chain.*

The lemma follows from the fact that the variables satisfy the equality

$$P(\xi_i \in E \mid \xi_j) P(\xi_k \in N \mid \xi_j) = P(\xi_i \in E, \ \xi_k \in N \mid \xi_j), \tag{5}$$

for any sets $E \in S_{x_i}$ and $N \in S_{x_k}$.

THEOREM 1. *Consider an interaction information $J(ijk)$ concerning the three random variables ξ_i, ξ_j, and ξ_k. Suppose that some two variables of them are mutually independent. Then $J(ijk) = 0$ occurs only when they form a simple Markov chain such that the two independent variables are the first and the third elements of the chain.*

Proof. Assume that the variables ξ_i and ξ_k are mutually independent. From Definition 1, $J(ijk) = EI(i, k \mid j)$. Thus, by Lemma 1, the fact that $J(ijk) = 0$ is equivalent to the Markovian property of ξ_i, ξ_j, and ξ_k in this order. Since an interaction information is symmetric with respect to each variable, the assertion of the theorem is clear from this. Q.E.D.

COROLLARY. *Let ξ_i, ξ_j, and ξ_k be connected by Markov dependence. Then $J(ijk) = 0$ implies the mutual independence of some two variables.*

The previous result is generalized to higher dimensional interaction informations. First, using the general formula of Kolmogorov and Dobrushin, we can make the next general definition with the analogous idea to McGill [4].

DEFINITION 2. A n-dimensional interaction information is given with regard to n random variables $\xi_{j_1}, \xi_{j_2}, \ldots, \xi_{j_n}$, where $(j_1, j_2, \ldots, j_n)$ is any permutation of n variable parameters,

$$\begin{aligned} J(j_1 j_2, \ldots, j_n) = {} & EI(j_1, j_n \mid (j_2, \ldots, j_{n-1})) \\ & - \Sigma' EI(j_1, j_n \mid (i_1', i_2', \ldots, i_{n-3}')) \\ & + \Sigma'' EI(j_1, j_n \mid (i_1'', i_2'', \ldots, i_{n-4}'')) \\ & - \cdots + (-1)^{n-1} \sum_{k=2}^{n-1} EI(j_1, j_n \mid j_k) + (-1)^n I(j_1, j_n), \end{aligned} \tag{6}$$

where $\sum'$, $\sum''$,..., denote the sum of the conditional informations under all admissible combinations $(i_1', i_2', \ldots, i_{n-3}')$, $(i_1'', i_2'', \ldots, i_{n-4}''), \ldots$, taken from the parameter set $\{j_2, j_3, \ldots, j_{n-1}\}$.

In general, since an interaction information is symmetric with respect to each variable, the above definition does not lose its generality. Moreover, clearly, it implies that Definition 1 as the case for $n = 3$.

THEOREM 2. *If the variables $\{\xi_j\}$ $(j = 1, 2, \ldots)$ are connected by Markov dependence, then n-dimensional interaction information of any separately selected n random variables $\xi_{j_1}, \xi_{j_2}, \ldots, \xi_{j_n}$, where $(j_1, j_2, \ldots, j_n)$ is any integer sequence in increasing order, takes negative and positive values corresponding to the dimensional degree to be odd and even, respectively, provided $n \geqslant 3$. Especially, under the same condition, its value goes to zero if and only if the variables ξ_{j_1} and ξ_{j_n} are mutually independent.*

Proof. As is well known [2], if the variables $\{\xi_{j_1}, \xi_{j_2}, \ldots, \xi_{j_n}\}$ follow the Markov relation, we can get

$$EI(j_1, j_n \mid \mathscr{P}) = 0,$$

where $\mathscr{P}$ is any subcombination, taken from $\{j_2, j_3, \ldots, j_{n-1}\}$. Thus,

$$J(j_1 j_2 \cdots j_n) = (-1)^n I(j_1, j_n). \tag{7}$$

This shows that each assertion of the theorem is true. Q.E.D.

3. INTERACTION INFORMATION OF THE HOMOGENEOUS MARKOV CHAIN

Now, consider the interaction information of the homogeneous Markov chain. As the dependence becomes a little stronger, the criterion of the interaction information zero relates with somewhat strict requirements.

THEOREM 3. *Let $\{\xi_j\}$ $(j = 1, 2, \ldots)$ be a simple, homogeneous Markov chain, taking finitely many states with the number m. Suppose that these states are only recurrent ones, and the associated transition matrix is given as*

$$\mathbf{P} = (p_{ij}), \qquad p_{ij} > 0, \qquad \sum_{j=1}^{m} p_{ij} = 1, \qquad \text{for each } i \quad (1 \leqslant i \leqslant m).$$

Then, for $m \geqslant n \geqslant 3$, next three conditions are equivalent. (a) *n-dimensional interaction information $J(j_1 j_2 \cdots j_n)$ of some, arbitrarily selected n different variables $\{\xi_{j_1}, \xi_{j_2}, \ldots, \xi_{j_n}\}$ is zero, where we assume $1 \leqslant j_1 < j_2 < \cdots < j_n$.*

(b) *Markov chain is $(n-2)$-dependent.*

(c) *The reduced matrix Q satisfies $Q^{n-1} = 0$, where*

$$Q = (q_{ij}), \qquad q_{ij} = p_{ij} - p_{mj} \quad (i, j = 1, 2, \ldots, m-1).$$

Proof. First prove (a) $\Leftrightarrow$ (b) and then (b) $\Leftrightarrow$ (c). Let (a) be satisfied. Then from (7),

$$0 = J(j_1 j_2 \cdots j_n) = (-1)^n I(j_1, j_n).$$

This implies by the homogeneity of the process, that our chain is $(j_n - j_1 - 1)$-dependent. However, since $j_n - j_1 - 1 \geqslant n - 2$, it requires at least, $(n-2)$-dependence. Thus, (a) $\Rightarrow$ (b). On the contrary, (b) $\Rightarrow$ (a) is clear. Next, (b) $\Leftrightarrow$ (c) will be proved.

Let $\mathbf{P}^{n-1}$ be given

$$\mathbf{P}^{n-1} = \begin{pmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_m \\ \vdots & \vdots & & \vdots \\ \lambda_1 & \lambda_2 & \cdots & \lambda_m \end{pmatrix} \lambda_i > 0, \qquad \sum_{i=1}^{m} \lambda_i = 1. \tag{8}$$

Then it requires that

$$\lambda_j = \sum_{k=1}^{m} p_{ik}^{(n-2)} p_{kj} = \sum_{k=1}^{m} p_{mk}^{(n-2)} p_{kj} \qquad \text{for any } i \text{ and } j, \ (1 \leqslant i, j \leqslant m-1) \tag{9}$$

where $p_{ik}^{(n-2)}$ is the conditional probability of the state k, given i, after $(n-2)$ transitions. However, if we apply the relation

$$p_{im}^{(n-2)} = 1 - \sum_{k=1}^{m-1} p_{ik}^{(n-2)} \qquad (i = 1, 2, \ldots, m),$$

we have

$$\sum_{k=1}^{m-1} (p_{ik}^{(n-2)} - p_{mk}^{(n-2)})(p_{kj} - p_{mj}) = 0 \qquad (1 \leqslant i, j \leqslant m-1). \tag{10}$$

On the other side, taking care that

$$p_{ik}^{(n-2)} - p_{mk}^{(n-2)} = \sum_{l=1}^{m} p_{il}^{(n-3)} p_{lk} - \sum_{l=1}^{m} p_{ml}^{(n-3)} p_{lk}$$
$$= \sum_{l=1}^{m-1} (p_{il}^{(n-3)} - p_{ml}^{(n-3)})(p_{lk} - p_{mk}),$$

we can obtain from this

$$\sum_{k=1}^{m-1} \sum_{l=1}^{m-1} (p_{il}^{(n-3)} - p_{ml}^{(n-3)})(p_{lk} - p_{mk})(p_{kj} - p_{mj}) = 0,$$

which is really a reduced form for n. Therefore, continuing this inductive operations, we can reach

$$\sum_{k_1=1}^{m-1} \sum_{k_2=1}^{m-1} \cdots \sum_{k_{n-2}=1}^{m-1} (p_{ik_1} - p_{mk_1})(p_{k_1k_2} - p_{mk_2}) \cdots (p_{k_{n-2}j} - p_{mj}) = 0, \tag{11}$$

i.e.,

$$\sum_{k_1=1}^{m-1} \sum_{k_2=1}^{m-1} \cdots \sum_{k_{n-2}=1}^{m-1} q_{ik_1} q_{k_1k_2} \cdots q_{k_{n-2}j} = q_{ij}^{(n-1)} = 0 \qquad (1 \leqslant i, j \leqslant m-1).$$

This means that $\mathbf{Q}^{n-1} = \mathbf{O}$. Thus, (b) $\Rightarrow$ (c). Conversely, assume that (c) holds. We know that a nilpotent matrix Q_1 of $(n-1)$th order can always be represented by some arbitrary regular matrix R

$$R^{-1}Q_1R = \begin{pmatrix} 0 & & \\ \vdots & \ddots & \\ 0 & \cdots & 0 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 0 & \cdots & 0 \\ & \ddots & \vdots \\ & & 0 \end{pmatrix}$$

where the matrix given on the right is a special triangular $(n-1) \times (n-1)$ matrix of which elements on the diagonal line and under this line (or, above this line) are all zero. Thus, since the state number is m, $m \geqslant n \geqslant 3$, the derived Markov chain from $\mathbf{Q}$ must be at least $(n-2)$-dependent. Q.E.D.

Remark. For $n > m \geqslant 2$, we can replace (b) by (b′).

(b′) Markov chain is $(m-2)$-dependent.

THEOREM 4. *If there exists a finite integer n such that for $m \geqslant n \geqslant 1$,*

$$\mathbf{P}^n = \begin{pmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_m \\ \vdots & \vdots & & \vdots \\ \lambda_1 & \lambda_2 & & \lambda_m \end{pmatrix} \lambda_i > 0, \qquad \sum_{i=1}^{m} \lambda_i = 1, \tag{12}$$

the distribution $(\lambda_1, \lambda_2, \ldots, \lambda_m)$ *is the absolute stationary probability distribution of the Markov chain with a transition matrix* $\mathbf{P}$. *It is, in fact, a* $(n-1)$*-dependent Markov chain of which state distribution becomes stationary after the first n-transitions, provided n is the smallest integer satisfying* (12).

Proof. Assume that the Markov chain has an absolute stationary distribution $(p_1, p_2, \ldots, p_m)$. Then,

$$p_j = \sum_{i=1}^{m} p_i p_{ij}, \qquad p_i > 0, \qquad \sum_{i=1}^{m} p_i = 1,$$

where all states are assumed to be recurrent. Thus, inductively,

$$p_j = \sum_{i=1}^{m} p_i p_{ij}^{(n)}.$$

Since $p_{ij}^{(n)} = \lambda_j$ $(i, j = 1, 2, \ldots, m)$, we have

$$p_j = \sum_{i=1}^{m} p_i \lambda_j = \lambda_j \qquad (1 \leqslant j \leqslant m)$$

Therefore, by the uniqueness of the absolute stationary distribution, λ_j values must be equivalent to those of p_j. The $(n-1)$-dependence is clear. Q.E.D.

COROLLARY. *Under the same conditions as Theorem* 3, *we have*

$$\lim_{|j_n - j_1| \to \infty} J(j_1 j_2 \cdots j_n) = 0.$$

This assertion is clear by the asymptotically independent property of such a Markov chain.

4. Interaction Information of the Gaussian Markov Chain

We show the decomposition theorem for Gaussian Markov chain using the interaction information zero criterion.

Assume that the mean values of the variables are all zero and any submatrices of the correlation coefficients are regular. Denote

$$\rho_{ij} = E\xi_i \xi_j \qquad (i, j = 1, 2, \ldots).$$

Using Lemma 1, we can get Lemma 2.

LEMMA 2. *For the Gaussian variables* ξ_1, ξ_2, *and* ξ_3, *they constitute a simple Markov chain in this order if and only if there exists a relation*

$$\rho_{12}\rho_{23} = \rho_{13}\rho_{22} \qquad \textit{provided} \quad \rho_{ii} \neq 0 \quad (i = 1, 2, 3). \tag{13}$$

THEOREM 5. *Let* $\{\xi_j\}$ $(j = 1, 2, \ldots)$ *be Gaussian variables, connected by Markov dependence. Choose some variables* $\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_n}$ *and let* $(i_1, i_2, \ldots, i_n)$ *be fixed. Then, in order to make the n-dimensional interaction information* $J(i_1 i_2 \cdots i_n)$ *be zero, it is necessary and sufficient that there exists at least one variable* ξ_k, $k \in [\alpha, \beta)$ *such that* ξ_k *and* ξ_{k+1} *are mutually independent, where*

$$\alpha = \min_{1 \leqslant k \leqslant n} i_k, \qquad \beta = \max_{1 \leqslant k \leqslant n} i_k,$$

and $[\alpha, \beta)$ *denotes a semiclosed interval consisting of all integers k such that* $\alpha \leqslant k < \beta$.

Proof. First of all, we show that it is sufficient to prove this for the successive random variables $\xi_{j_1}, \xi_{j_2}, \ldots, \xi_{j_N} (\alpha = j_1 < j_2 < \cdots < j_N = \beta)$. In fact, under the Markovian relation, each requirement of

$$J(i_1 i_2 \cdots i_n) = 0$$

and

$$J(j_1 j_2 \cdots j_N) = 0 \tag{14}$$

is the same by Theorem 2. Clearly, for the establishment of (14), the variables ξ_{j_1} and ξ_{j_N} must be mutually independent.

Now, since the sufficiency of the theorem is clear, we prove its necessity. Assume that for any s, $(s = 1, 2, \ldots, N-1)$

$$\rho_{j_s j_{s+1}} \neq 0. \tag{15}$$

Then by (13),

$$\rho_{j_{s-1} j_{s+1}} \neq 0 \qquad s = 2, 3, \ldots, N-1.$$

This procedure can be developed for the variables ξ_{j_r}, $\xi_{j_{r+1}}$ and $\xi_{j_{s+1}}$ inductively, and we obtain

$$\rho_{j_r j_{s+1}} \neq 0, \qquad r = 1, 2, \ldots, s-1.$$

Just as the same way, for ξ_{j_1}, ξ_{j_t}, and $\xi_{j_{t+1}}$,

$$\rho_{j_1 j_{t+1}} \neq 0 \qquad t = s, s+1, \ldots, N-1.$$

This implies $\rho_{j_1 j_N} \neq 0$, so that (15) is a contradiction. Q.E.D.

This theorem will be more effective if we consider it with Theorem 2.

Acknowledgment

The author sincerely expresses his many thanks to Professor K. Kunisawa for his kind advice during the completion of this paper.

References

[1] Dobrushin, R. L. (1959). General formulations of the basic theorems of Shannon in the information theory. *Russian Math. Surveys.* **XIV**, 6, 3–104.

[2] Doob, J. L. (1953). *Stochostic Processes.* John Wiley and Sons, New York/London.

[3] Kullback, S. (1959). *Information Theory and Statistics.* John Wiley and Sons, New York/London.

[4] McGill, W. J. (1954). Multivariate information transmission. *I.R.E. Trans. Information Theory* 93–111. (1954 Symposium on Information theory)

[5] Pinsker, M. S. (1964). *Information and Information Stability of Random Variables and Processes.* Academy of Sciences, USSR, Moscow; Holden-Day, San Francisco/London/Amsterdam (English).

[6] Sakaguchi, M. (1967). Interaction Information in multivariate probability distributions. *Kōdai Math. Semi. Rep.* **19**, 147–155.

[7] Gelfand, I. M. and Yaglom, A. M. (1957). Calculation of the amount of information about a random function contained in another such function. *Russian Math. Surveys* **XII**, 3, 3–52.